

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- (i) APPLICANT: Yechezkel Kashi et al.
- (ii) TITLE OF INVENTION: ABUNDANT, WELL DISTRIBUTED AND
HYPERPOLYMORPHIC SIMPLE SEQUENCE REPEATS
IN PROKARYOTE GENOMES AND USE OF SAME FOR
PROKARYOTE CLASSIFICATION AND TYPING
- (iii) NUMBER OF SEQUENCES: 42
- (iv) CORRESPONDENCE ADDRESS:
 - (A) ADDRESSEE: Mark M. Friedman c/o Anthony Castorina
 - (B) STREET: 2001 Jefferson Davis Highway, Suite 207
 - (C) CITY: Arlington
 - (D) STATE: Virginia
 - (E) COUNTRY: United States of America
 - (F) ZIP: 22202
- (v) COMPUTER READABLE FORM:
 - (A) MEDIUM TYPE: 1.44 megabyte, 3.5" microdisk
 - (B) COMPUTER: Twinhead® Slimnote-890TX
 - (C) OPERATING SYSTEM: MS DOS version 6.2,
Windows version 3.11
 - (D) SOFTWARE: Word for Windows version 2.0 converted to
an ASCII file
- (vi) CURRENT APPLICATION DATA:
 - (A) APPLICATION NUMBER:
 - (B) FILING DATE:
 - (C) CLASSIFICATION:
- (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER:
 - (B) FILING DATE:
- (viii) ATTORNEY/AGENT INFORMATION:
 - (A) NAME: Friedmam, Mark M.
 - (B) REGISTRATION NUMBER: 33,883
 - (C) REFERENCE/DOCKET NUMBER: 910/8
- (ix) TELECOMMUNICATION INFORMATION:
 - (A) TELEPHONE: 972-3-5625553
 - (B) TELEFAX: 972-3-5625554
 - (C) TELEX:

(2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 23
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

GATTTTGCAT ATGAGTATAT TAC 23

(2) INFORMATION FOR SEQ ID NO:2:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 21
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

TTAATTACAG GATGTTTCAGT C 21

(2) INFORMATION FOR SEQ ID NO:2:

- (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:
AATTTATCCG GTGAATGTGG T 21

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:
CAACTTAATC TCGGGCTGAC 20

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:
TACATGGCTG ATTATGCGG 20

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:
TCGCTATGAA TATCTACTGA C 21

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:
GTCAGAGCAG ATCCAGAATG 20

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:
TCTACAGCAA ATGAACAATG 20

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:
GGTCATCAGG TGAAATAATC 20

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:
CGTCCTGATA GATAAAGTGT C 21

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:
CAATGGAAC TACCAATGAC 20

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:
TACCGCGAAG AATTCAACAC 20

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:
AGCATCAGCG CACAATGCAC 20

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:
TGTATGCAGG-CTGGCACAAC 20

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:
ATAACGATCT CCATATCTAC 20

(2) INFORMATION FOR SEQ ID NO:16:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 20
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:
CTCTATCAGC AACTTCTGCC 20
- (2) INFORMATION FOR SEQ ID NO:17:
- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 20
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:
ATCCGCAGGA TTTTCGCACC 20
- (2) INFORMATION FOR SEQ ID NO:18:
- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 20
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:
TGCCAGCGTA AATCCGCAAC 20
- (2) INFORMATION FOR SEQ ID NO:19:
- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 2
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:
GT 2
- (2) INFORMATION FOR SEQ ID NO:20:
- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 2
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:
TC 2
- (2) INFORMATION FOR SEQ ID NO:21:
- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 3
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:
CGG 3
- (2) INFORMATION FOR SEQ ID NO:22:
- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 3
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double

- (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

GGT 3

- (2) INFORMATION FOR SEQ ID NO:23:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 4
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

ATTA 4

- (2) INFORMATION FOR SEQ ID NO:24:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 4
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

CTGG 4

- (2) INFORMATION FOR SEQ ID NO:25:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 259
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

```

GTTATGTCCT ATCCACGGT ATTAAATATG GTTCATTAGG ATGTTTATTT 50
CTTGATTTTG CATATGAGTA TATTACCCCC CCTCAAAAA AATAAATTAA 100
TTAAATGAT GGCTTATATA AAATAAAATT TAAAGCAAGG AATCTCAATG 150
GATGTTAAAC AAAATGAGAT TTTGTGAAAG CAATAAATTA TTGACTTCGT 200
TTTAGATTGG TTTAGCTATA ATGTTATACA TTCAAATGAC TGAACATCCT 250
GTAATTAAA                                     259

```

- (2) INFORMATION FOR SEQ ID NO:26:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 256
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

```

TTGTTATGTC TTATCCACG GATTTAATA TAGTTCATTT GGATGTCAT 50
TTCTTTATTT TGCATATGAG TATATTACCC CTCAAAAAA TAAATTAATT 100
AAAACGATTG CTTATATAAA ACAAAATTTA AAGCAAGGAA TCTCAATGGA 150
TGTTAAACAA AATGAGATTT AGTGAAACA ATAAATTATT CACTTCGTTT 200
TAGATTTGTT TAGCTATAAT GTTATACATT CAAATGACTG AACATCCTGT 250
ATTAA                                     256

```

- (2) INFORMATION FOR SEQ ID NO:27:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 257
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

```

TTTGTTATGT CTTATCCAC GGTATTAAAT ATAGTTCATT TGGATGTTCA 50

```

TTTCTTTATT TTGCATATGA GTATATTACC CCTTCAAAAA ATAAATTAAT 100
TAAAACGATT GCTTATATAA AACAAAAATT AAAGCAAGGA ATCTCAATGG 150
ATGTTAAACA AAATGAGATT TAGTGAAAAC AATAAATTAT TCACTTCGTT 200
TTAGATTGTG TTAGCTATAA TGTATACAT TCAATGACT GAACATCCTG 250
TAATTAA 257

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 263
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

TCTATGTTCT TATCNCACG GTXTTAAATA TGGTTCATTA GGATGTTTAT 50
TTCTTGATT TTGCATATGAG TATATTACCC CCCCCCTCA AAAAAATAAA 100
TTAATTAATAA TGATGGCTTA TATXAAATAA AATTTAAAGC AAGGAXTCTC 150
AATGGATGTT AAACAXAATG AGATTTTGTG AAXGCXATXX ATTATTGXCT 200
TCGTTGTAXA TTTGCTXAGC TATAATGTTA TXCATTCAA TGACTGAACA 250
TCCTGTXXTT AXA 263

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 264
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

XXXXTXXXX XTXXCCAC GGTATTTAAT ATXGTTTATT XGGATGTTXA 50
TTTCTTXATT TTGCATATGA GTATATTAXX XXXXCCCTX XAAAAATAA 100
ATTAATTAATA AXGATXGCTT ATATAAAXA AAATTTAAAG CAAGGAATCT 150
CAATGGATGT TAAACAAAAT GAGATTTXGT GAAAXCAATA AATTATTXAC 200
TTCGTTXTAG ATTTGXTTAG CTATAATGTT ATACATTCAA ATGACTGAAC 250
ATCCTGTAAT TAAX 264

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 181
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

TTTXXXXCGA AAAAAATAGG AAAGGGGGGG GGGCTAATCG GCAGGGAAGG 50
CCGCCCCGGA TAGCGGGCGG CAXAAGGAAT CAXAATTTC AGGTCAGACG 100
GGCTGCAAGT TGCAGACCGT TAAATCATC GGXXGGGGTG TCGTACCACA 150
CTTTACCTGC CGTCAGCCCG AGATTAAGTT G 181

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 181
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

TTTXXXXCGG AAAAAATXG AAAGGGGGGG GGGCTAATCG GCAGGGAAGG 50
CCGCCCCGGA TAGCGGGCGG CAGAAGGAAT CAGAATTTC AGGTCAGACG 100
GGCTGCAAGT TGCAGACCGT TAAATCATC GGTGGGGGTG TCGTACCACA 150
CTTTACCTGC CGTCAGCCCG AGATTAAGTT G 181

(2) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 177
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

```
TTTXCCGGAA AAAAATXGAA AGGGGGGGGC TAATCGGCAG GGAAGGCCGC 50
CCCGGATAGC GGGCGGCAGA AGGAATCAGA ATTTCCAGGT CAGATGGGCT 100
GCAAGTTGCA GACCGTTATA ATCATCGGTT GGGGTGTCGT ACCACACTTT 150
ACCTGCCGTC AGCCCGAGAT TAAGTTG 177
```

(2) INFORMATION FOR SEQ ID NO:33:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 182
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

```
TXTXCGGAAA AAAAXAGGAA AGGGGGGGGG GCTAATCGGC AGGGAAGGCC 50
GCCCCGGATA GCGGGCGGCA GAAGGAATCA GAATTTCAG GTCCAGACGGG 100
CTGCAAGTTG CAGACCGTTA AAATCATCGG TTGGGGTGTC GTACCACACT 150
TTACCTGCCG TCAGCCCGAG ATTAAAGTTT GG 182
```

(2) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 175
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

```
XCGGAAAAAA ATGAAAGGGG GGGGCTAAT CGGCAGGGAA GGCCGCCCCG 50
GATAGCGGGC GGCAGAAGGA ATCAGAATT CCAGGTCAGA TGGGCTGCAA 100
GTTGCAGACC GTTATAATCA TCGGTTGGGG TGTCGTACCA CACTTTACCT 150
GCCGTCAGCC CGAGATAAAG TTTGG 175
```

(2) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 174
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

```
CGGAAAAAAA TGAAAGGGG GGGGCTAATC GGCAGGGAAG GCCGCCCCGG 50
ATAGCGGGCG GCAGAAGGAA TCAGAATTTC CAGGTCAGAT GGGCTGCAAG 100
TTGCAGACCG TTATAATCAT CGGTGGGGT GTCGTACCAC ACTTTACCTG 150
CCGTCAGCCG GAGATAAAGT TTGG 174
```

(2) INFORMATION FOR SEQ ID NO:36:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 182
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

```
XXXXXXXXXX XAAAAAXXG AAAXXGGGGG GGGCTAATCG GCAGGGAAGG 50
CCGCCCGGA TAGCGGGCGG CAGAAGGAAT CAGAATTTC AGGTCAGAXG 100
GGCTGCAAGT TGCAGACCGT TAXAATCATC GGTGGGGTG TCGTACCACA 150
```

CTTTACCTGC CGTCAGCCCG AGATXAAGTT XG

182

(2) INFORMATION FOR SEQ ID NO:37:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 8
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

GGGGGGGG 8

(2) INFORMATION FOR SEQ ID NO:38:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 9
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

GGGGGGGGG 9

(2) INFORMATION FOR SEQ ID NO:39:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

GGGGGGGGGG 10

(2) INFORMATION FOR SEQ ID NO:40:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 4
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

TAAA 4

(2) INFORMATION FOR SEQ ID NO:41:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 4
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

TTAA 4

(2) INFORMATION FOR SEQ ID NO:42:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

TTAAA 5